Research article

The dot-probe attention bias task as a method to assess psychological wellbeing after anesthesia: A study with adult female long-tailed macaques (Macaca fascicularis)

Lauren C. Cassidy[a,b], Emily J. Bethell[c,d], Ralf R. Brockhausen[a], Susann Boretius[b,e], Stefan Treue[a,b], and Dana Pfefferle[a,b]

[a] Welfare and Cognition Group, Cognitive Neuroscience Laboratory, German Primate Center - Leibniz Institute for Primate Research, Goettingen, Germany

[b] Leibniz-ScienceCampus Primate Cognition, German Primate Center & University of Goettingen, Goettingen, Germany

[c] Liverpool John Moores University, Research Centre in Evolutionary Anthropology and Palaeoecology, Liverpool, England

[d] Liverpool John Moores University, Research Centre in Brain and Behaviour, Liverpool, England

[e] Functional Imaging Laboratory, German Primate Center - Leibniz Institute for Primate Research, Goettingen, Germany

Short title: Dot-probe as a measure of animal psychological wellbeing

Corresponding Author:

Lauren C. Cassidy

Welfare and Cognition Group, Cognitive Neuroscience Laboratory

German Primate Center - Leibniz Institute for Primate Research

Kellnerweg 4

Goettingen, 37085, Germany

+49 551 3851-361

lcassidy@dpz.eu

Number of Tables: 2

Number of Figures: 4

Word count: Please indicate the word count including Abstract and body text. This is not to include the title page, reference list or figure legends.

Keywords: dot-probe task; attention bias; affect; emotion; anesthesia

1

Abstract

Understanding the impact routine research and laboratory procedures have on animals is crucial to improving their wellbeing and to the success and reproducibility of the research they are involved in. Cognitive measures of welfare offer insight into animals' internal psychological state, but require validation. Attention bias - the tendency to attend to one type of information over another – is a cognitive phenomenon documented in humans and animals that is known to be modulated by affective state (i.e., emotions). Hence, changes in attention bias may offer researchers a deeper perspective of their animals' psychological wellbeing. The dot-probe task is an established method for quantifying attention bias in humans (by measuring reaction time to a dot-probe replacing pairs of stimuli), but has yet to be validated in animals. We developed a dot-probe task for long-tailed macaques (*Macaca fascicularis*) to determine if the task can detect changes in attention bias following anesthesia, a context known to modulate attention and trigger physiological arousal in macaques. Our task included the following features: stimulus pairs of threatening and neutral facial expressions of conspecifics and their scrambled counterparts, two stimuli durations (100 and 1000 ms), and counterbalancing of the dot-probe's position on the touchscreen (left, right) and location relative to the threatening stimulus. We tested eight group-housed adult females on different days relative to being anesthetized (baseline and one-, three-, seven-, and 14-days after). At baseline, monkeys were vigilant to threatening content when stimulus pairs were presented for 100 ms, but not 1000 ms. On the day immediately following anesthesia, we found evidence that attention bias changed to an avoidance of threatening content. Attention bias returned to threat vigilance by the third day post-anesthesia and remained so up to the last day of testing (14 days after anesthesia). We also found that attention bias was independent of the type of stimuli pair (i.e., whole face vs. scrambled counterparts), suggesting that the scrambled stimuli retained aspects of the original stimuli. Nevertheless, whole faces were more salient to the monkeys as responses to these trials were generally slower than to scrambled stimulus pairs. Overall, our study suggests it is feasible to detect changes in attention bias following anesthesia using the dot-probe task in non-human primates. Our results also reveal important aspects of stimulus preparation and experimental design.

Introduction

Ensuring high standards of animal welfare is crucial for conducting ethical and reproducible biomedical and basic research. Ideally methods for assessing welfare should be objective and reflect changes to an animal's physiological or/and psychological wellbeing. Attention bias, the process of selectively attending to one type of information over another [1], is one cognitive process that may offer insight into animals' psychological wellbeing and affective state. Human attention biases are influenced by context, changes in physiology, mood, and intrinsic traits such as personality [2–4]. Attention bias tasks have found that humans, particularly for those with affective disorders like anxiety, preferentially attend to threatening information [5–9]. Given this evidence, tasks for detecting affect-mediated attention biases are being modified for animals to assess affect non-invasively [10].

Affect-mediated attention bias tasks are adapted for animals using biologically relevant stimuli that trigger innate responses such as gaze (e.g., [11]) or movement (e.g., [12]; reviewed in [10,13]). Differences or changes in attention biases have been examined via trait affect (parrots, *Amazona amazonica*: [14]; rhesus macaques, *Macaca mulatta*: [15]), by manipulating state affect in individual animals (rhesus macaques: [16]) or groups of animals (mice, *Mus musculus*: [12]; starlings, *Sturnus vulgaris* : [17,18]), and by comparing groups of animals administered with or without pharmacological anxiety drugs (cattle, *Bos taurus*: [19]; sheep, *Ovis aries*: [20]). For example, Bethell et al. [16] found that how male rhesus macaques attended to threatening and neutral facial expressions was modulated by the type of affect manipulation the males recently experienced. Specifically, males were more avoidant of threatening stimuli following a stressful veterinary procedure (i.e., health check) than after period of enrichment.

Looking-time experiments can help provide a complete picture of the different processing stages of attention: initial engagement, maintenance, disengagement [21–23]. However, these experiments can be time consuming (e.g., if video must be coded) or costly (if eye-tracking equipment is required) [24,25]. One alternative to looking-time experiments is the dot-probe task, which is sensitive to affect-mediated attention bias in humans (reviewed in [26–28]). In this task, participants are presented with a stimulus pair (e.g., facial expressions) simultaneously for a fixed duration. After the stimuli disappear, a 'dot-probe' (i.e., neutral target) appears in the location of one stimulus and the latency to touch this target is measured. Faster reaction times to the dot-probe indicate that attention was likely allocated towards the stimulus it replaced, whereas slower reactions suggest that attention shifted from another location, presumably the other stimulus. Manipulating the stimuli presentation duration allows researchers to capture the different stages of attention [29],

3

which may reveal if participants show vigilance, avoidance, or a pattern of both to a particular stimulus [30,31]. Stimulus pairs often consist of a stimulus with neutral content paired with another of high threatening content, with the latter capturing gaze automatically. Importantly – and different to other tasks measuring attention bias – these stimuli are task irrelevant (i.e., no trained reward contingencies) and may limit habituation due to their biological salience for the species being studied. Furthermore, animals can learn the dot-probe task easily as touching the dot-probe is the only rule they have to understand (e.g., [32]).

So far the detection of affect-mediated attention biases by the dot-probe task has been tested only in humans. A meta-analysis of studies investigating anxiety found that anxious participants were faster to react to dot-probes replacing the negative or threatening stimulus [26]. Similar findings have been reported for humans suffering from depression [27]. These findings attest that the dot-probe task is sensitive to trait affect and have provided the foundation for dot-probe studies testing context driven attention changes. In this respect, dot-probe studies in humans involving affective manipulations have tested negatively-valenced contexts ranging from acute stress induction (cold press test: [33]; mild contextual shock: [34]) to putatively severe, chronic stressors (rocket attack: [35]; combat deployment: [36]). How and if attention bias is modulated as detected by the dot-probe task may depend on gender (e.g., [33]) and level of stress exposure (e.g., [34–36]).

Given the supporting evidence from human studies, the dot-probe task shows potential for detecting affect-driven attention bias changes in other animals. Despite this potential, the dot-probe task has been implemented relatively rarely within the realm of animal cognition (reviewed in [10,28]). Currently, dot-probe studies have been conducted only in non-human primates (NHPs), focusing on comparing reactions to dot-probes replacing affective content to those replacing neutral content (bonobos, *Pan paniscus*: [32]; chimpanzees, *Pan troglodytes*: [38]; rhesus macaques: [42]; capuchins, *Sapujus apella*: [43]; summarized in Supp. Table 5). Yet no study to date has tested whether the task is also sensitive to changes in the affective state in these species, which bears potential as a welfare assessment method.

General anesthesia is a common and necessary procedure in veterinary medicine. Experiencing anesthesia is likely one of the strongest contexts that could influence affect in captivity, as it is a known physiological stressor (e.g., [44–46]). In addition to the anesthesia itself, associated processes, such as social group separation for fasting, having the anesthetic applied, and waking up from surgery in isolation, likely cause additional physiological or/and psychological effects. We opportunistically tested the reliability and sensitivity of a dot-probe task for detecting changes in affect due to experiencing prolonged anesthesia in eight adult female long-tailed macaques (*Macaca*

4

fascicularis). Improving methodologies for assessing NHP psychological wellbeing is necessary as these species are crucial for the advancement of scientific and medical knowledge, treatments, and applications (reviewed in [47–49]). We tested the monkeys on the dot-probe task during a baseline test session, when no anesthesia had been administered at least 30 days prior, and at four test sessions following prolonged anesthesia (one-, three-, seven-, and 14-days). Our dot-probe task for NHPs incorporated design features common among dot-probe studies (summarized in [28]). For this experiment, we assessed whether the dot-probe task detected **(Q1)** attention bias, **(Q2)** an affect-mediated change in attention bias following the anesthesia, and **(Q3)** when attention bias returned to baseline levels post-anesthesia assuming a change occurred. We expected the dot-probe task to detect an attention bias to threat (specifically reacting more quickly to dot-probes replacing the aggressive face) for whole face stimuli during the baseline test session. Additionally, we predicted that the dot-probe task would be able to detect a change in attention bias following anesthesia that would return to the monkeys' baseline levels of attention bias in the days following. We present our study as a guide for optimizing future studies as it is the first to implement the dot-probe in relation to an affect manipulation in an animal.

Materials and Methods

Study subjects and housing facility

We conducted the study on eight adult female long-tailed macaques living at the German Primate Center, Goettingen, Germany. The monkeys were housed in isosexual groups of four to five individuals with visual and auditory contact to other macaque groups. Age of the monkeys ranged from 6 to 19 years (mean ± standard deviation: 11.3 ± 5.8 years) at the time point of the study. Housing consisted of a large indoor compartment with a 12-hour light/dark cycle (from 07:00 to 19:00) connected by an elevated tunnel to an outdoor compartment where animals could experience natural light, temperature fluctuations, and wind, with visual access to the outdoors (living space exceeded the size requirements for macaques set by EU directive 2010/63/EU). Both areas were furnished with fixed and dynamic perching (e.g., raised platforms, ropes), environmental enrichment (e.g., balls, cardboard), and carpeted with wood shavings. A flexible compartment (i.e., testing compartment) adjacent to the indoor living quarters was used for animal training, testing, temporary separation, and veterinary procedures (approximately 80 cm by 75 cm by 90 cm). Monkeys had access to water and monkey chow *ad libitum*, and received fresh fruit and vegetables daily.

Experimental testing protocol

5

Our study took place between August 2017 and January 2018 and ran concurrently with another project investigating the effect of prolonged anesthesia on the brain using magnetic resonance imaging (MRI; see Statement of ethics for permit information). Veterinarians regularly monitored the monkeys during the entire study. In preparation for anesthesia, monkeys were separated (but with visual, acoustic, and olfactory contact to their group members) and food removed the night before. Anesthesia was induced by a mixture of ketamine (mean ± standard deviation: 8.0 ± 2.7 mg per kg) and medetomidine (mean ± standard deviation: 0.02 ± 0.01 mg per kg) and maintained by isoflurane (0.8 to 1.7% in oxygen and ambient air) via an endotracheal tube and pressure controlled active ventilation. The duration of isoflurane anesthesia ranged from 213 to 350 min (310 ± 42 min).

Following anesthesia, monkeys were kept separate overnight (but with visual, acoustic, and olfactory contact to their group members) for the purposes of recovery and observation. Monkeys were returned to their living quarters the following morning. Each monkey performed the dot-probe task once as a baseline, when no anesthesia had been administered at least 27 days prior, and at four time points following the anesthesia session: on average one day (A + 1d), three days (A + 3d), seven days (A + 7d), and 14 days (A + 14d) after. Baseline measurements occurred in a counterbalanced design. Five monkeys were tested at least 27 days before any monkeys in the group experienced prolonged anesthesia (range: 28 to 32 days). Three monkeys were tested at least 34 days after all anesthesia procedures occurred (range: 35 to 36 days). Due to the timing of baseline test sessions, we presume that these test sessions coincided with a period of comparatively low stress to the sessions immediately following prolonged anesthesia. It is possible daily fluctuations in stress may have occurred on the day of each test session due to social, environmental, or/and husbandry factors for example. However, these influences are likely to be limited as we observed no increases in aggression (rare occurrence overall) or changes in hierarchy that may have indicated group instability (systematic behavioral observations did occur, but were not the focus of this study). Additionally, our veterinarians did not observe any indications of poor health throughout the study.

Dot-probe tests occurred between 12:00 and 15:30 and ran until the monkey completed the task (described in Dot-probe task). We controlled for touchscreen apparatus and location tested within individuals (two monkeys were tested in their indoor living quarters and five in the adjacent testing compartment with visual, acoustic, and olfactory contact to their group members). Differences in testing location were driven by organizational issues and did not induce differences in overall reaction time (reaction times of the two monkeys performing the test in their living quarters was within the range of monkeys performing the test in the testing compartments).

Cognitive task

6

Experimental testing apparatus and software

We used a cage-based touchscreen system (XBI) developed within the Cognitive Neuroscience Laboratory at the German Primate Center to administer the dot-probe task to the monkeys [50,51]. The XBI is a computerized training and testing system that allows NHPs to learn and carry out complex cognitive tasks with no physical restraints and minimal supervision. Fluid reward (i.e., juice) can be delivered via a reward tube positioned at a fixed distance of 24.0 cm from the touchscreen (30.4 cm by 22.8 cm in size; [50]). Dot-probe tasks for training and testing purposes were programmed and carried out in MWorks (version 0.6; http://mworksproject.org/), a highly flexible open-source C++ -based package for designing and real-time controlling behavioral tasks [50].

Dot-probe task

We based our dot-probe task on design features common in human and NHP studies. These features included:

(1) Paired affective face stimuli (with aggressive and neutral expressions) known to capture attention bias (i.e., whole face stimuli).

(2) Scrambled counterparts of the whole face stimuli, where facial expressions were not recognizable (i.e., scrambled stimuli). Scrambled stimuli are important to include as they help validate the salience of the whole face stimuli for the species being studied. Importantly, these stimuli verify that attention biases are attributable to the affective content present in whole face stimuli, rather than low-level stimuli features (e.g., luminance; see Stimuli Preparation).

(3) Dot-probes replacing the aggressive (congruent trial) or neutral (incongruent trial) stimulus to detect attention bias and if changes occurring in relation to prolonged anesthesia.

(4) Two stimuli durations of 100 ms and 1000 ms to differentiate the stages of early and late attention processing. Including multiple stimuli durations could reveal which stimulus monkeys initially orient towards (100 ms) and if they maintain or disengage attention from that stimulus at a later processing stage (1000 ms).

(5) Counterbalancing of the position of the dot-probe and aggressive stimulus with respects to congruency and location on touchscreen (left, right) for each stimulus pair. Counterbalancing controlled for possible effects due to hemispheric laterality, where the right brain hemisphere processes affective stimuli faster [52–54].

7

The time course of the dot-probe task is depicted and described in Fig. 1. Further details about the positioning and size of the stimuli, dot-probe, and trial structure of the dot-probe task can be found in the supplementary material (see section "Dot-probe task design").

INSERT Fig. 1

Task training protocol

All monkeys were naïve towards using the touchscreen system and receiving diluted juice reward (50 % grape juice, 50 % water) as a positive reinforcer. Training stimuli preparation and training steps for the dot-probe task are described in detail in supplementary material (see sections "Training stimuli" and "Dot-probe training task and procedure"). Monkeys were considered trained when they were able to successfully complete (correctly initiate a trial and touch the dot-probe) 80 % of trials initiated during a training session (usually lasting 20 min).

Stimuli preparation

Stimuli presented during the test sessions consisted of whole faces, scrambled versions of the whole face stimuli, and filler stimuli (grey filler, scrambled filler, social filler; [55]). Filler stimuli were included as warm-up trials and to separate blocks in the dot-probe task. Grey filler stimuli were pairs of 11.0 cm by 11.0 cm grey squares (RGB values: 191, 191, 191) included as warm-up trials. We selected scrambled filler stimulus pairs (N = 3) from the training image set that marked the beginning of each block (one per block); each pair consisted of mirrored scrambled images of an unknown adult conspecific male with a neutral expression (3 actors). Social filler stimulus pairs (N = 9) each consisted of mirrored square images of an unknown conspecific infant (9 actors) and were included at the end of each block of the dot-probe task (three per block). Whole face stimuli were of 'original' and morphed images of unknown conspecific adult males (i.e., actors) with neutral (N = 18) and aggressive (N = 18) expressions paired by actor identity (i.e., stimulus pairs). Aggressive expressions were characterized by raised eyebrows, intent staring, or/and open-mouths with teeth often visible (e.g., [56]). In contrast, neutral expressions were characterized by a closed mouth and relaxed face (e.g., [56]).

We processed all stimuli using Adobe Photoshop CS3 Extended (version 10.0). Generally, images were cropped around the head to remove background content. Additionally, we "closed" the eyes of each neutral image to minimize potential affective content as macaques perceive eye contact as threatening [57], and eyes are salient facial substructures for these species [58]. Aggressive and neutral stimulus pairs (matched by actor identity) were adjusted to be similar in color and luminance. Values for color and luminance were measured using the histogram function in Adobe Photoshop.

8

Aggressive and neutral stimuli did not differ in color or luminance as determined by using Linear Mixed Models (LMMs), with facial expression as the only predictor and actor identity as a random effect (red: χ^2 = 1.05, df = 1, p = 0.305; green: χ^2 = 0.18, df = 1, p = 0.675; blue: χ^2 = 0.71, df = 1, p = 0.400; luminance: χ^2 = 0.18, df = 1, p = 0.669). Furthermore, aggressive and neutral images did not differ in volume (i.e., number of pixels; volume: χ^2 = 1.496, df = 1, p = 0.221).

After processing, each stimulus was formatted on a grey square background. Morphed images were created by merging two formatted 'original' stimuli of the same facial expression using FantaMorph software (version 5.4.8) and were processed similarly. Once all whole face stimuli were formatted, we scrambled 10 by 10-pixel squares within the shape of the head for all images in MATLAB (version 9.0.0.341360). The position of whole face and scrambled stimulus pairs were counterbalanced by side and displayed to face towards the center of the touchscreen. See supplementary material section "Stimuli preparation and additional analytics" for more detailed information regarding all stimuli preparation and analytics.

Data preparation

Our response variable was latency to touch the dot-probe (i.e., reaction time) replacing whole face or scrambled stimulus pairs. Altogether the monkeys correctly completed (touched the dot-probe within the 10 s presentation window) 2608 100 ms trials and 2535 1000 ms trials (whole face and scrambled). We treated our response data similar to other touchscreen studies with NHPs (e.g., [37,40,59]). Accordingly, we removed reaction time responses less than 350 ms (100 ms trials: N = 1, 0 %; 1000 ms trials: N = 72, 2.8 %) as they were deemed too quick to reflect a meaningful response to the stimuli ([31]; see also [60]). Nearly two-thirds (63.9 %) of the removed trials (N = 47 trials) came from one monkey (monkey E). We removed this monkey's data from the analysis of 1000 ms trials (N = 185; 8.7 % of selected data) since it was likely she was anticipating the position of the dot-probe during these trials. Similar to [32], we checked slow responses (those over 1200 ms) by video to assess whether they reflected a true response or recording error, due to the touchscreen not registering the first touch of the dot-probe for example. Slow responses occurred most often due to the monkey adopting an unconventional position (e.g., sitting between the touchscreen and reward tube), or being distracted (100 ms trials removed: N = 175; 1000 ms removed: N = 380; [40]). Lastly, we excluded trials two standard deviations above the mean of responses grouped by each animal, time period, trial type, dot-probe position, and congruency to reduce the influence of outliers (100 ms trials: N = 117, 4.5 %; 1000 ms trials: N = 97; 3.8 %; [32,34,37]). Overall, we entered 2365 100 ms trials (90.7 % of original data set) and 1932 1000 ms trials (76.2 % of original data set) into our statistical analyses.

9

Statistical analyses

Our statistical analysis explored if the dot-probe task could detect **(Q1)** a general attention bias towards threat in our study population, **(Q2)** an affect-mediated change in attention bias in relation to prolonged anesthesia (i.e., MRI procedure), and **(Q3)** when attention bias returned to baseline levels following prolonged anesthesia (if a change was found). We tested these questions by fitting Linear Mixed Models (LMMs) to our data that was transformed to meet the assumptions of normally distributed residuals. While the use of non-normal distributions with reaction time data is encouraged (e.g., [61]), we opted for transformations and LMMs due to model convergence difficulties and to reduce the impact of outliers, which occurred due to the task's 10 s response window for the dot-probe [62].

The data were split by stimuli duration to differentiate responses due to different stages of attention processing [29]. **Q1:** To determine if our task detected a general attention bias towards threat during the baseline test session, we tested if the monkeys' reaction times to the dot-probe were influenced by the interaction of trial type and congruency. Here we predicted that reaction time to congruent trials (dot-probe replacing the aggressive stimulus) would be faster than incongruent trials (dot-probe replacing the neutral stimulus) for whole face stimulus pairs, but not scrambled stimulus pairs. We also included dot-probe position into the interaction structure of these models to control for well documented hemispheric lateralization effects (reviewed in [53,54]). **Q2:** To test if attention bias changed due to prolonged anesthesia, we fit an LMM using data from the baseline session and the session immediately following prolonged anesthesia (A + 1d), including test session as an interacting factor. We predicted that monkeys changed how they responded to the task during the session immediately following prolonged anesthesia in comparison to baseline responses, either by becoming slower to respond to congruent trials or faster to respond to incongruent trials. As is common practice in the literature and to investigate attention bias effects further, we fit an additional LMM examining the effect of test session on the mean attention bias score (AB score) for each individual per session using the transformed data. The attention bias scores were calculated by subtracting the monkeys' mean reaction time to congruent trials from their mean reaction time to congruent trials per test session. **Q3:** To determine when attention bias effects returned to baseline, we used data of the baseline session and sessions three-, seven-, and 14-days following prolonged anesthesia. The model we fit was informed by the final predictor structure of question two, including test session, trial type, congruency, and their interactions where relevant. We also controlled for dot-probe position, the time testing began, trial number, and rank (Table 1) when feasible as these are known factors influencing attention bias [15].

10

INSERT Table 1

We conducted our statistical analyses in R (version 3.6.1; [63]) and fit LMMs using the 'lme4' package (version 1.1-21; [64]). Data were transformed using the 'powerTransform' function within the 'car' package (version 3.0-2; [65]). This function estimated a lambda value (λ) based on the response distribution, which we used to conduct a Box-Cox transformation (((y^λ)-1)/λ) of reaction time for each model. Additionally, we checked the distributions of test and control predictors for normality and z-transformed covariates to a mean of zero and a standard deviation of one to provide more comparable estimates and aid the interpretation of interactions [66,67]. All test and control predictors were also checked for correlations (none above 0.5). Monkey identity was included as a random effect and with all possible random slopes to keep type I error rates at the nominal level of 0.05 [68,69]. Correlations between random slopes and intercepts were excluded in the models to allow model convergence and decrease computation time as their exclusion does not compromise type I error rates [69].

Each LMM was fit using the function 'lmer' with a Gaussian error structure and identity link function [70] using Maximum Likelihood [71]. To aid model convergence for reaction time models, we used the argument 'control' to specify the 'bobyqa' optimizer and increased the number of iterations to 100,000. We checked the assumptions of normally distributed data and homogenous residuals by visually inspecting a qqplot and a scattered plot of the residuals plotted against the fitted values. The distributions of random effects were also checked for normality. Additionally, we assessed model stability by excluding subjects one at a time and comparing the subset model estimates with those from the full data set. Variance Inflation Factors (VIF) were derived using the 'vif' function within the 'car' package (version 3.0-2; [65]) from a standard linear model lacking random effects and including all predictor and control variables separately (no interaction term) to rule out collinearity [72–75]. All model checks indicated no deviations from the assumptions of a normally distributed model. Furthermore, model stability checks indicated that there were no influential cases and collinearity was ruled out as an issue (maximum VIF = 1.28).

To determine if the test predictors had a meaningful effect on reaction time to the dot-probe replacing the stimuli, we used a likelihood ratio test (LRT) comparing each model to its null counterpart (LMM with the intercept specified at 1 and lacking all test predictors) using the 'anova' function with the argument 'test' set to "Chisq" [76,77]. Our significance criterion was set to consider interactions and predictors with p-values under 0.100. We examined the significance of interacting test predictors by using the 'drop1' function, which compares the full model to respective reduced models in a series of LRTs [69]; we removed those interactions and their respective interacting

11

random slopes that did not have an effect (although the main effects were retained). The drop1 function was used to determine *p*-values for the remaining interactions and main effects once the final model was deduced. Lastly, we examined the robustness of the final models (where relevant) using the following measures: effect size as measured by the function 'r.squaredGLMM' from the 'MuMIn' package [version 1.43.17; [78]]; repeatability, the proportion of variation attributed to between-subject or group variation, was calculated using the 'rpt' from the 'rptR' (version 0.9.22; [79,80]); test predictor power was tested using the 'powerSim' function of the "simr" package using their respective estimate in the final model and set to 1000 simulations and nominal confidence of 0.05 (version 1.0.5; [81]).

As six of eight monkeys were left-handed, we conducted a post-hoc analysis using the data and original models of question one (i.e., does the task detect attention bias?) to check if where the dot-probe was positioned in relation the monkeys' preferred hand (i.e., ipsilateral or contralateral to preferred hand) had more explanatory power than dot-probe position (i.e., left or right side of touchscreen), which may have masked any effect of hemispheric lateralization in our study. Therefore, we coded each trial to reflect the location of the dot-probe in relation to the preferred hand of the individual being tested (i.e., position preferred hand: ipsilateral or contralateral; Table 1). The post-hoc analysis consisted of an information theoretic approach to evaluate the goodness-of-fit, Akaike Information Criterion (AICc) scores, differences in AICc scores (ΔAICc), and Akaike weights (scales models relative to one another) of the two models of interest (one for each stimuli duration; [82,83]). Specifically, we compared the full model including position preferred hand (replacing the variable dot-probe position) to the original full model for each stimuli duration using the 'aictab' function of the 'AICcmodavg' package (version 2.2.2; [84]), which ranks the models based on the selected Akaike information criteria.

Results

All eight monkeys reached the 80 % performance training criterion to be considered to participate in the dot-probe experiment. Seven monkeys participated in all five sessions that the dot-probe task was administered; monkey B refused to participate in the first session following anesthesia (A + 1d), but participated in the other four sessions. Each monkey was able to finish the dot-probe task for each session she participated in (exposed to 144 whole face and scrambled trials per session), except in one instance during A + 1d where it was necessary to stop the task and return the monkey to the home cage due to the cage being required for unrelated veterinary purposes (monkey E during A + 1d, exposed to 119 whole face and scrambled trials). Additionally, monkey E's third test session following prolonged anesthesia occurred after eight days and was categorized the A + 7d test

12

session. Monkey A's first test session immediately after experiencing prolonged anesthesia occurred two days after and was grouped with the A + 1d data. Mean reaction time across all testing days to the dot-probe replacing whole face and scrambled stimuli for 100 ms trials was 668 ± 252 ms and 709 ± 263 ms for 1000 ms trials (see Supp. Table 6 for more detailed information).

Q1: Do we find attention bias during baseline? – Attention bias towards threatening content at 100 ms (but not 1000 ms).

Our first question addressed whether the expected pattern of attention bias towards threat was evident in our study sample at baseline. If so, we predicted faster responses to probes replacing aggressive (congruent) versus neutral (incongruent) stimuli.

The model comparison between the full and null model for 100 ms stimuli duration was significant (LRT: χ^2 = 16.41, df = 6, p = 0.012). The final model did not include any interactions between congruency, trial type, and dot-probe position, but did indicate significant main effects of each of these variables (Table 2; raw data for each predictor are plotted per monkey in the supplementary material section "Supplementary figures of raw data"). As predicted, monkeys were faster to respond to congruent trials than incongruent trials (LRT: χ^2 = 4.04, df = 1, p = 0.045; Table 2; Fig. 2) demonstrating an attention bias towards stimuli with threatening content at baseline. A significant effect of trial type indicated monkeys responded slower to dot-probes replacing whole face stimuli as compared to scrambled stimuli (LRT: χ^2 = 7.39, df = 1, p = 0.007; Table 2; Fig. 2). However, congruency and trial type did not interact significantly, indicating the overall pattern of attention bias towards threatening content did not differ between the whole face and scrambled stimuli. The final model had a fixed effect variance (marginal R^2) of 0.29, a fixed and random effects variance (conditional R^2) of 0.53, and repeatability measurement of 0.18. Predictor variable power was 90.7 % for trial type (confidence interval range: 88.7 % - 92.4 %) and 58.1 % for congruency (confidence interval range: 55.0 % - 61.2 %).

The model comparison between the full and null model for 1000 ms stimuli duration trials was non-significant (LRT: χ^2 = 3.56, df = 7, p = 0.829).

INSERT Fig. 2

Q2: Do we see affect mediated changes in attention bias following prolonged anesthesia? – Attention bias switches to avoidance of threatening content.

Our second question focused on whether attention bias changed following prolonged anesthesia. If so, we predicted an interaction between test session (baseline versus A + 1d) and congruency.

13

The model comparison between the full and null model for 100 ms stimuli duration trials was significant (LRT: χ^2 = 13.95, df = 5, p = 0.012). The final model included interactions between test session and congruency (LRT: χ^2 = 3.461, df = 1, p = 0.063; Table 2), and test session and trial type (LRT: χ^2 = 4.643, df = 1, p = 0.031; Table 2). On the day immediately following anesthesia, monkeys were slower to respond to congruent trials than incongruent trials, with the opposite pattern seen at baseline (Fig. 3; raw data are plotted by test session per monkey in the supplementary material section "Supplementary figures of raw data"). This result suggests prolonged anesthesia triggered a change in how monkeys responded to threatening information following prolonged anesthesia compared to baseline. Additionally, while monkeys were generally slower to respond to whole face trials than scrambled trials during the baseline session, that difference diminished on the day immediately after anesthesia. The final model had a fixed effect variance (marginal R^2) of 0.09, a fixed and random effects variance (conditional R^2) of 0.47, and repeatability measurement of 0.33. Predictor variable power was 58.7 % for the interaction of test session and trial type (confidence interval range: 55.6 % - 61.8 %) and 48.3 % for the interaction of test session congruency (confidence interval range: 45.7 % - 51.5 %).

Following common practice in the literature and to explore the interaction of test session and congruency further, we compared attention bias difference scores between baseline and A +1d to assess the direction of change. The model comparison between the full and null model for the attention bias difference scores was significant (LRT: χ^2 = 3.66, df = 1, p = 0.056). Overall, monkeys had a positive attention bias score during the baseline period, and a negative attention bias score on the day immediately following prolonged anesthesia (A + 1d; Supp. Table 7; Fig. 3). This change in attention bias score suggests that monkeys were vigilant to threatening content (positive attention bias score) during the baseline test session and more avoidant of threat (negative attention bias score) on the day immediately following anesthesia. The attention bias score model had a fixed effect variance (marginal R^2) of 0.26 and a fixed and random effects variance (conditional R^2) of 0.66. Repeatability could not be calculated for the attention bias score model, likely due to the low number of observations in the model. Predictor variable power was 55.1 % for test session (confidence interval range: 52.0 % - 58.2 %).

INSERT Fig. 3

Q3: When does attention bias return to baseline levels post-anesthesia? – Attention bias recovered to baseline levels by three days following prolonged anesthesia.

Our third question investigated when attention bias recovered to baseline levels following prolonged anesthesia. The model comparison between the full and null model for 100 ms stimuli duration trials

14

during the baseline test session was significant (LRT: χ^2 = 28.60, df = 11, p = 0.003). The final model did not include any interactions found in the second analysis (test session and congruency, test session and trial type), but did indicate significant main effects of test session, congruency, and trial type (Table 2). Overall, monkeys were faster to respond to congruent trials than incongruent trials (LRT: χ^2 = 7.71, df = 1, p = 0.005; Table 2; Fig. 4; raw data are plotted by test session per monkey in the supplementary material section "Supplementary figures of raw data"). This finding suggests that attention bias recovered to baseline levels by the third day post-anesthesia. There was also an effect of trial type, where monkeys were slower to respond to dot-probes replacing whole face stimuli than to those replacing scrambled stimuli (LRT: χ^2 = 5.48, df = 1, p = 0.019; Table 2). Lastly, reaction time differed between test sessions, where monkeys became faster to respond to dot-probes each session in the two weeks following anesthesia (LRT: χ^2 = 8.43, df = 3, p = 0.038; Table 2; Fig. 4). The final model had a fixed effect variance (marginal R^2) of 0.09, a fixed and random effects variance (conditional R^2) of 0.62, and repeatability measurement of 0.12. Predictor variable power was 76.90 % for trial type (confidence interval range: 74.7 % - 79.5 %) and 83.0 % for congruency (confidence interval range: 80.5 % - 85.3 %).

INSERT Table 2

INSERT Fig. 4

Hand-preference influences response time to dot-probes

Across several of the reaction time analyses, we noticed that dot-probe position had a significant effect. That is, during the baseline test session (LRT: χ^2 = 4.11, df = 1, p = 0.043; Table 2; Supp. Fig. 2), as well as on three-, seven-, and 14-days post-anesthesia (LRT: χ^2 = 3.58, df = 1, p = 0.058; Table 2), monkeys also responded to dot-probes positioned on the left faster than those positioned on the right.

Due to these effects and our observation that most monkeys were left-handed (6 of 8 monkeys), we checked if where the dot-probe appeared in relation to the monkeys' preferred hand (i.e., position preferred hand) explained the data better than the dot-probe position (left or right on touchscreen) for both stimuli durations. The post-hoc information theoretic analysis for the 100 ms stimuli duration indicated the best model was the original model including dot-probe position over the one including the variable position preferred hand, with an Akaike weight of 0.95 (of 1.00 of the Akaike model weights combined). This analysis provides evidence that the output of the original model including dot-probe position is warranted for stimulus pairs presented for 100 ms (Supp. Table 4; [82]). In contrast, the model including position preferred hand for the 1000 ms stimuli duration was

15

the best model in comparison to the original model including dot-probe position, with an Akaike weight of 0.99, indicating evidence as the best model of the two (Supp. Table 4; [82]).

Discussion

Validating techniques for assessing animal welfare is essential for determining which tools are the most informative, sensitive, and reliable. Such endeavors will help researchers focus on welfare indices that are most useful and thereby enhance the scientific outcomes of projects involving animal research models [85]. Well-established affect-mediated attention bias methods from human cognitive psychology research show promise for use in other animals. Before being used to assess welfare, these tasks must be tested using a context known to change other indices of welfare (e.g., physiological responses). Therefore, we assessed the potential of the dot-probe task for detecting psychological changes, as measured by attention biases, after prolonged anesthesia in adult female long-tailed macaques.

Similar to other dot-probe studies in humans (e.g., [33–36]), we found that the monkeys, when not stressed, were biased towards threatening stimuli (versus neutral) in the dot-probe task. This attention bias effect was only found when stimuli were presented for the shorter duration (i.e., 100 ms) and not for the longer duration (i.e., 1000 ms). Likely the 100 ms stimuli duration captured information about where the monkeys' attention was automatically drawn (voluntary eye-saccades in NHP species occur 241 ms on average; [86]). Conversely, the 1000 ms stimuli duration may have encouraged the monkeys to anticipate where the dot-probe was going to appear by reaching towards or/and touching stimuli before they disappeared. One monkey (monkey E), for example, frequently touched stimuli shown for 1000 ms before the dot-probe appeared. Human and other NHP dot-probe studies have also found attention bias to affective stimuli for durations ranging from 40 ms to 1000 ms [26,32,39,40]. In contrast, studies in chimpanzees and sub-adult rhesus macaques have not found attention biases to affective stimuli when presented for 33 ms, 150 ms, 300 ms, and 500 ms [37,38,42]. Stimuli in the chimpanzee studies, however, were presented in greyscale, which may hamper affect recognition [87]. While our findings for the 100 ms stimuli duration are new to the NHP dot-probe literature (duration has not been tested before), the lack of attention bias effects for the 1000 ms stimuli duration contradicts other NHP studies [39,40,43]. Presently it is difficult to tease apart whether these differences in findings are species specific or due to variance in the type of stimuli or/and duration presented. Based on our findings, attention biases in NHPs may be best captured using briefly presented color stimuli. Looking-time measures would further guide the selection of stimuli durations in future studies.

16

We also found evidence that the cognitive and affective load of stimuli influenced how monkeys responded. During the baseline test session, monkeys responded slower to dot-probes following stimuli of whole faces than scrambled faces. Processing faces is known to be more cognitively demanding (e.g., [88]) and can potentially trigger affective reactions in macaques (e.g., [59,60,89]). Such influences might account for the response-slowing we observed.

Unexpectedly, our data show an attention bias towards the threatening stimulus even if stimulus pairs were scrambled. It seems that our scrambled stimuli retained affective features of the original image despite our efforts to equate threatening and neutral images for basic image features (see next paragraph for an in-depth discussion). Currently, the inclusion of scrambled stimuli is lacking in dot-probe experiments involving NHPs as only one other study has previously done so (but see [38]). While stimuli need to be carefully designed, scrambled images should also be included to rule out the influence of unforeseen low-level features.

So why did responses to scrambled and whole face stimuli not differ significantly? This effect was most likely due to our scrambling methodology retaining the face shape of the original image. For example, our aggressive faces tended to have a higher height to width ratio due to face lengthening caused by the dropped jaw of an open-mouth threat face. Such basic shape cues of facially expressed affect allow for an ultra-fast initial estimate of an affective state as these are likely amongst the fastest steps of initial cortical image processing. In conditions of high noise, such shape cues might in fact be the only information available. Given that our scrambling did not remove characteristics of face shape, these cues might be sufficient to cause an affect-based response and therefore diminish the differences in responses between our type of scrambled and whole face stimuli. Importantly, however, the lack of a difference between responses to our scrambled and whole face stimuli does not preclude that the dot-probe effect is based on the affective content of stimuli. We provide specific guidance on scrambling images for future studies in the recommendations section below.

As anesthesia is known to trigger changes in physiological indices of welfare and attention bias in macaques (e.g., [16,44–46]), we sought to use it as a way to validate the dot-probe task as a measurement of psychological welfare. We found that on the day immediately following prolonged anesthesia the monkeys' pattern of attention bias deviated from baseline levels on the day immediately following prolonged anesthesia. Specifically, attention bias changed from a vigilance to threat to an avoidance of threat. This pattern appeared despite any typical post-surgical side effects, such as general response slowing or other general motor function deficiencies and, hence, reflects changes in psychological wellbeing. By the third day after anesthesia, the monkeys' pattern of attention bias (and thus their psychological wellbeing) had returned to the baseline pattern.

17

Although our sample size is small and model robustness measures low, our findings match what has been found in relation to acute and chronic stressors using the dot-probe task in humans [33–36].

Unsurprisingly, monkeys generally responded faster to dot-probes when positioned ipsilateral to their preferred hand. This result confirms the importance of counterbalancing stimuli presentation by side, as implemented in our study and accounted for in our analyses. Hand-preference effects might occur in parallel to the hemispheric lateralization of affective stimuli, a subtle process that can influence such kinds of behavioral asymmetries in animals (reviewed in [53,54]).

Inter-individual traits likely play an influential role in how monkeys attend to facial stimuli, respond to, and recover from stressors (e.g., [15]). These differences may result in varying levels of attention bias to facial stimuli and changes to attention bias following stressors that are more or less pronounced. One monkey in our study (monkey F), for example, showed a particularly strong pattern of avoidance and seemed to be the most agitated (e.g., pacing, hesitant to enter testing compartment) on the day immediately following prolonged anesthesia (personal observation). Monkeys that exhibit such patterns of avoidance may be more affected by prolonged anesthesia than others, and could be of greatest concern from a welfare perspective. Previous work from our lab also found that avoidance or behavioral inhibition to touch threatening stimuli was positively correlated with fearful temperament scores on a human intruder task [59]. Our present data support the suggestion that such individuals may need to be the focus for animal care staff following procedures of putatively heightened stress.

Overall, our results indicate that the dot-probe task can detect affect-mediated changes in attention bias using colored stimuli presented for 100 ms in adult long-tailed macaque females. We hope that others will use this proof-of-concept study as a framework to improve upon our attention bias task for assessing procedure severity. While studies with small sample sizes are problematic, these numbers are the reality for the type of procedures (i.e., prolonged anesthesia) that are typically performed in NHP research. With more studies we can improve severity assessment measures further and build a fuller picture of the impact invasive procedures have on animal psychological wellbeing. In the following sections, we provide several recommendations for improving the dot-probe task and experimental design considering our results.

Recommendations for setting up a dot-probe experiment with animals

How to maximize attention bias effects

Attention biases are best detected when the difference between the attributes of interest – such as affective intensity of stimulus pairs – is maximized (e.g., [5]). Therefore, careful stimuli selection is

18

necessary to maximize attention bias effects. In our case, we maximized the difference in affective intensity between aggressive and neutral images by selecting whole face images of adult male conspecifics. Adult male long-tailed macaques are considered to be the most threatening individuals for females due to their larger canines, body size, and high dominance status [90,91]. Furthermore, we edited the eyes of neutral images to appear closed as direct eye contact can also be threatening to macaques [57]. Alternatively, the difference in affective intensity can be maximized when threatening images are paired with those where gaze is averted (e.g., [60]). Using whole body images, which can contain postural information about affective state, rather than faces could also maximize attention bias effects (e.g., [32,92]).

How to prepare facial stimuli

Once stimuli have been selected, adjusting stimulus pairs (and the population of images as a whole) to be similar in color, luminance, and size is standard practice for NHP dot-probe studies (e.g., [37,38,41]). Cropping out background content scenery of task stimuli is another encouraged, but inconsistently applied practice (e.g., [38,42]). Unless specific precautions are taken, cropping may retain affective features such as head shape or body posture (when the whole body is shown). Since these features may particularly affect the neutrality of scrambled stimuli, applying a visual mask of standardized dimensions to stimuli (e.g., [93]) before scrambling is the most effective method to avoid this issue. Such masks should also match the color of the touchscreen so that images of interest are the only aspects in the task capturing attention.

Scrambled stimuli should reflect whole face stimulus pairs for all factors except for those of interest. Including these stimuli can authenticate the salience of the whole face stimuli and that attention bias effects are due to the removed (affective) content. Incorporating and examining the effects of scrambled stimuli into future studies will enhance the validity and reliability of the dot-probe task as a welfare assessment tool in animals.

How to organize trials

Task trial structure may affect dot-probe task responses as effects elicited from whole face stimuli may carry over to scrambled stimuli trials if they are interleaved. Affect studies in humans have found that responses to stimuli can be modulated by trial structure, specifically if they are interleaved within the same block or in a separate block of trials (blocked design; e.g., [94,95]). The aim of the trial structure in our study was to minimize cumulative arousal (e.g., [96,97]). Future dot-probe studies should consider testing blocks containing only one type of stimuli if cumulative arousal remains low.

> *How to minimize response variation*
>
> Variation in response times might be due to fluctuating changes in motivation over the course of the experiment (e.g., satiation), anticipatory responding, or distraction from the task. These sources of variation can be reduced by making some adjustments to task design and programming. For instance, motivation could be improved using a variable reward schedule (e.g., [98]) or limiting the time the animal has to respond to the dot-probe. Variation due to anticipatory behavior can be avoided by programming trials to abort if a stimulus (e.g., aggressive face) is touched before the dot-probe appears. Distraction (e.g., due to sounds from the surrounding environment) might be limited by training animals to hold on to a fixation cue throughout the presentation of the stimulus pairs (e.g., [37]).

Summary

Scientific understanding of animal affect and its overlap with psychological wellbeing is a burgeoning field, whose methods for detecting changes to these states require reliability and validation. Our study is the first to document the sensitivity of the dot-probe task for detecting changes in the affective state of non-human primates after prolonged anesthesia. Our data show that prolonged anesthesia did change the monkeys' affective state, which recovered to the baseline state by the third day after the procedure. In addition, we emphasize the importance of careful task stimuli selection and preparation, the inclusion of scrambled stimuli, trial structure, and the minimization of response variation. Future studies should explore the use of the dot-probe in other contexts that elicit varying valences of affect (e.g., positive, weakly negative) to understand the breadth of animals' affective niche [10,99]. Together, these investigations will form a complete picture of animal affect, will advise strategies for minimizing sources of stress, and enhance captive animal wellbeing. Such investigations are critical for understanding which experimental procedures and husbandry practices enhance welfare and promote positive affective states, ultimately improving ethical standards of practice and reproducible research.

Acknowledgments

Statement of ethics

20

Research with non-human primates (NHPs) represents a small but indispensable component of basic and biomedical research. The researchers of this study are aware and are committed to the great responsibility they have in ensuring the best possible science with the least possible harm to the monkeys [47]. The monkeys in this study were group-housed in the facilities of the German Primate Center (DPZ) in Goettingen, Germany. The facility provides the monkeys with an enriched environment including a multitude of toys and wooden structures, natural as well as artificial light and exceeding the size requirements of the European regulations, including access to outdoor space. The monkeys' psychological and veterinary welfare was monitored by the DPZ's staff veterinarians, the animal facility staff and the lab's researchers, all specialized in working with NHPs.

All aspects of this study complied with the German Primate Center guidelines for monkey care and use and were conducted in accordance with the German Animal Protection Law and the European Union Directive 2010/63/EU on the Protection of Animals used for Scientific Purposes. Monkeys used in this study were highly accustomed to their testing environment and voluntarily participated in the dot-probe task. Furthermore, the task was positively reinforced with juice and can be considered cognitive enrichment. The concurrent anesthesia study was approved by the responsible regional government office (Animal Welfare Service, Lower Saxony State Office for Consumer Protection and Food Safety (LAVES), license-number 33.19-42502-04-16/2278).

Conflict of interest statement

Funding sources

Author contributions

LC, EB, DP, and RB contributed to the conception and design of the study. RB programmed the dot-probe task. SB provided access to the monkeys and conducted the anesthesia study. ST provided the test systems and infrastructure support. LC trained the monkeys and collected the data. LC performed statistical analyses, with advice from EB and DP. LC, EB, and DP interpreted the data. LC drafted the manuscript, with substantial revision by DP, EB, and ST. All authors read, revised, and approved the submitted version of the manuscript.

Data availability statement

The data set can be accessed on the Goettingen Research Online website: https://doi.org/10.25625/ATAAGM (Reviewers: Data will become publicly available upon publication). Code for the dot-probe task can be provided on request.

References

1 MacLeod C, Mathews A, Tata P. Attentional bias in emotional disorders. J Abnorm Psychol. 1986 Feb;95(1):15–20.

2 Paul ES, Harding EJ, Mendl M. Measuring emotional processes in animals: The utility of a cognitive approach. Neurosci Biobehav Rev. 2005 May;29(3):469–91.

3 Désiré L, Boissy A, Veissier I. Emotions in farm animals: A new approach to animal welfare in applied ethology. Behav Processes. 2002 Nov;60(2):165–80.

4 Clore GL, Palmer J. Affective guidance of intelligent agents: How emotion controls cognition. Cognitive Systems Research. 2009 Mar;10(1):21–30.

5 Cisler JM, Koster EHW. Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. Clin Psychol Rev. 2010 Mar 1;30(2):203–16.

6 McNally RJ. Attentional bias for threat: Crisis or opportunity? Clin Psychol Rev. 2019 Apr;69:4–13.

7 Mogg K, Bradley BP. A cognitive-motivational analysis of anxiety. Behav Res Ther. 1998 Sep;36(9):809–48.

8 Barry TJ, Vervliet B, Hermans D. An integrative review of attention biases and their contribution to treatment for anxiety disorders. Front Psychol. 2015 Jul;6:1–5.

9 Veerapa E, Grandgenevre P, El Fayoumi M, Vinnac B, Haelewyn O, Szaffarczyk S, et al. Attentional bias towards negative stimuli in healthy individuals and the effects of trait anxiety. Sci Rep. 2020 Dec;10(1):11826.

10 Crump A, Arnott G, Bethell E. Affect-driven attention biases as animal welfare indicators: Review and methods. Animals. 2018 Aug 7;8(8):136.

11 Bethell E, Holmes A, MacLarnon A, Semple S. Cognitive bias in a non-human primate: Husbandry procedures influence cognitive indicators of psychological well-being in captive rhesus macaques. Anim Welf. 2012 May;21(2):185–95.

12 Trevarthen AC, Kappel S, Roberts C, Finnegan EM, Paul ES, Planas-Sitjà I, et al. Measuring affect-related cognitive bias: Do mice in opposite affective states react differently to negative and positive stimuli? Parker M, editor. PLoS One. 2019 Dec 30;14(12):e0226438.

13 Mendl M, Burman OHP, Parker RMA, Paul ES. Cognitive bias as an indicator of animal emotion and welfare: Emerging evidence and underlying mechanisms. Appl Anim Behav Sci. 2009 May;118(3-4):161–81.

14 Cussen VA, Mench JA. Personality predicts cognitive bias in captive psittacines, *Amazona amazonica*. Anim Behav. 2014 Mar;89:123–30.

15 Howarth ER, Kemp C, Thatcher H, Szott ID, Farningham D, Whitham CL, et al. Developing and validating attention bias tools for assessing trait and state affect in animals: A worked example with *Macaca mulatta*. Appl Anim Behav Sci. 2021 Jan;234:105198.

16 Bethell EJ, Holmes A, MacLarnon A, Semple S. Evidence that emotion mediates social attention in rhesus macaques. PLoS One. 2012 Aug;7(8):e44387.

17 Brilot BO, Asher L, Bateson M. Water bathing alters the speedaccuracy trade-off of escape flights in European starlings. Anim Behav. 2009 Oct;78(4):801–7.

18 Brilot BO, Bateson M. Water bathing alters threat perception in starlings. Biol Lett. 2012 Jun;8(3):379–81.

19 Lee C, Cafe LM, Robinson SL, Doyle RE, Lea JM, Small AH, et al. Anxiety influences attention bias but not flight speed and crush score in beef cattle. Appl Anim Behav Sci. 2018 Aug;205:210–5.

20 Lee C, Verbeek E, Doyle R, Bateson M. Attention bias to threat indicates anxiety differences in sheep. Biol Lett. 2016 Jun;12(6):20150977.

21 Posner MI. Orienting of attention. Q J Exp Psychol. 1980 Feb;32(1):3–25.

22 Posner MI, Petersen SE. The attention system of the human brain. Annu Rev Neurosci. 1990 Mar;13:25–42.

23 Petersen SE, Posner MI. The attention system of the human brain: 20 years after. Annu Rev Neurosci. 2012 Apr;35(1):73–89.

24 Winters S, Dubuc C, Higham JP. Perspectives: The looking time experimental paradigm in studies of animal visual perception and cognition. Hauber ME, editor. Ethology. 2015 Jul;121(7):625–40.

25 Hopper LM, Gulli RA, Howard LH, Kano F, Krupenye C, Ryan AM, et al. The application of noninvasive, restraint-free eye-tracking methods for use with nonhuman primates. Behavior Research Methods. 2020 Sep;53:1003–30.

26 Bar-Haim Y, Lamy D, Pergamin L, Bakermans-Kranenburg MJ, van IJzendoorn MH. Threat-related attentional bias in anxious and nonanxious individuals: A meta-analytic study. Psychol Bull. 2007 Feb;133(1):1–24.

27 Peckham AD, McHugh RK, Otto MW. A meta-analysis of the magnitude of biased attention in depression. Depress Anxiety. 2010 Dec;27(12):1135–42.

28 van Rooijen R, Ploeger A, Kret ME. The dot-probe task to measure emotional attention: A suitable measure in comparative studies? Psychon Bull Rev. 2017 Jan;24(6):1686–717.

29 Cooper RM, Langton SRH. Attentional bias to angry faces using the dot-probe task? It depends when you look for it. Behav Res Ther. 2006 Sep;44(9):1321–9.

30 Bradley BP, Mogg K, Falla SJ, Hamilton LR. Attentional bias for threatening facial expressions in anxiety: Manipulation of stimulus duration. Cogn Emot. 1998 Nov;12(6):737–53.

31 Koster EH, Crombez G, Verschuere B, Van Damme S, Wiersema JR. Components of attentional bias to threat in high trait anxiety: Facilitated engagement, impaired disengagement, and attentional avoidance. Behav Res Ther. 2006 Dec;44(12):1757–71.

32 Kret ME, Jaasma L, Bionda T, Wijnen JG. Bonobos (*Pan paniscus*) show an attentional bias toward conspecifics' emotions. Proc Natl Acad Sci. 2016 Apr;113(14):3761–6.

33 Carr AR, Scully A, Webb M, Felmingham KL. Gender differences in salivary alpha-amylase and attentional bias towards negative facial expressions following acute stress induction. Cogn Emot. 2016 Feb;30(2):315–24.

34 Shechner T, Pelc T, Pine DS, Fox NA, Bar-Haim Y. Flexible attention deployment in threatening contexts: An instructed fear conditioning study. Emotion. 2012 Oct;12(5):1041–9.

35 Wald I, Shechner T, Bitton S, Holoshitz Y, Charney DS, Muller D, et al. Attention bias away from threat during life threatening danger predicts PTSD symptoms at one-year follow-up. Depress Anxiety. 2011 May;28(5):406–11.

36 Sipos ML, Bar-Haim Y, Abend R, Adler AB, Bliese PD. Postdeployment threat-related attention bias interacts with combat exposure to account for PTSD and anxiety symptoms in soldiers. Depress Anxiety. 2014 Feb;31(2):124–9.

37 Wilson DA, Tomonaga M. Exploring attentional bias towards threatening faces in chimpanzees using the dot probe task. Chao L, editor. PLoS One. 2018 Nov;13(11):e0207378.

38 Kret ME, Muramatsu A, Matsuzawa T. Emotion processing across and within species: A comparison between humans (*Homo sapiens*) and chimpanzees (*Pan troglodytes*). J Comp Psychol. 2018 Nov;132(4):395–409.

39 King HM, Kurdziel LB, Meyer JS, Lacreuse A. Effects of testosterone on attention and memory for emotional stimuli in male rhesus monkeys. Psychoneuroendocrinology. 2012 Mar;37(3):396–409.

40 Lacreuse A, Schatz K, Strazzullo S, King HM, Ready R. Attentional biases and memory for emotional stimuli in men and male rhesus monkeys. Anim Cogn. 2013 Nov;16(6):861–71.

41 Parr LA, Modi M, Siebert E, Young LJ. Intranasal oxytocin selectively attenuates rhesus monkeys' attention to negative facial expressions. Psychoneuroendocrinology. 2013 Sep;38(9):1748–56.

42 Morin EL, Howell BR, Meyer JS, Sanchez MM. Effects of early maternal care on adolescent attention bias to threat in nonhuman primates. Dev Cogn Neurosci. 2019 Aug;38:100643.

43 Schino G, Carducci P, Truppa V. Attention to social stimuli is modulated by sex and exposure time in tufted capuchin monkeys. Anim Behav. 2020 Mar;161:39–47.

44 Lee VK, Flynt KS, Haag LM, Taylor DK. Comparison of the effects of ketamine, ketamine-medetomidine, and ketamine-midazolam on physiologic parameters and anesthesia-induced stress in rhesus (*Macaca mulatta*) and cynomolgus (*Macaca fascicularis*) macaques. J Am Assoc Lab Anim Sci. 2010 Jan;49(1):57–63.

45 Whitten PL, Stavisky R, Aureli F, Russell E. Response of fecal cortisol to stress in captive chimpanzees (*Pan troglodytes*). Am J Primatol. 1998 Jan;44(1):57–69.

46 Novak MA, Hamel AF, Kelly BJ, Dettmer AM, Meyer JS. Stress, the HPA axis, and nonhuman primate well-being: A review. Appl Anim Behav Sci. 2013 Jan;143(2):135–49.

47 Roelfsema PR, Treue S. Basic neuroscience research with nonhuman primates: A small but indispensable component of biomedical research. NeuroImage. 2014 Jun;82(6):1200–4.

48 Cox LA, Olivier M, Spradling-Reeves K, Karere GM, Comuzzie AG, VandeBerg JL. Nonhuman primates and translational research–cardiovascular disease. ILAR J. 2017 Dec;58(2):235–50.

49 Friedman H, Ator N, Haigwood N, Newsome W, Allan JS, Golos TG, et al. The critical role of nonhuman primates in medical research. Pathog Immun. 2017 Aug;2(3):352–65.

50 Calapai A, Berger M, Niessing M, Heisig K, Brockhausen R, Treue S, et al. A cage-based training, cognitive testing and enrichment system optimized for rhesus macaques in neuroscience research. Behav Res Methods. 2017 Feb;49(1):35–45.

51 Berger M, Calapai A, Stephan V, Niessing M, Burchardt L, Gail A, et al. Standardized automated training of rhesus monkeys for neuroscience research in their housing environment. J Neurophysiol. 2018 Mar;119(3):796–807.

52 Fox E. Processing emotional facial expressions: The role of anxiety and awareness. Cogn Affect Behav Neurosci. 2002 Mar;2(1):52–63.

53 Forrester GS, Todd BK. A comparative perspective on lateral biases and social behavior. In: Progress in Brain Research. First. London, UK: Academic Press; 2018. p. 377–403.

54 Leliveld LMC, Langbein J, Puppe B. The emergence of emotional lateralization: Evidence in non-human vertebrates and implications for farm animals. Appl Anim Behav Sci. 2013 Apr;145(1-2):1–4.

55 Pfefferle D. Facial Stimuli - Macaques [Internet]. figshare. Dataset; 2020. Available from: https://figshare.com/articles/dataset/Facial_Stimuli-_Macaques/13227779.

56 Maréchal L, Levy X, Meints K, Majolo B. Experience-based human perception of facial expressions in Barbary macaques (*Macaca sylvanus*). PeerJ. 2017 Jun;5:e3413.

57 van Hooff JA. The facial displays of the catarrhine monkeys and apes. In: Morris D, editor. Primate Ethology. London: Weidenfeld & Nicolson; 1967. p. 7–68.

58 Guo K, Robertson RG, Mahmoodi S, Tadmor Y, Young MP. How do monkeys view faces?a study of eye movements. Exp Brain Res. 2003 Jun;150(3):363–74.

59 Bethell EJ, Cassidy LC, Brockhausen RR, Pfefferle D. Toward a standardized test of fearful temperament in primates: A sensitive alternative to the human intruder task for laboratory-housed rhesus macaques (*Macaca mulatta*). Front Psychol. 2019 May;10:1–7.

60 Bethell E, Holmes A, MacLarnon A, Semple S. Emotion evaluation and response slowing in a non-human primate: New directions for cognitive bias measures of animal emotion? Behav Sci. 2016;6(1):2.

61 Lo S, Andrews S. To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. Front Psychol. 2015 Aug;6:1171.

62 Whelan R. Effective Analysis of Reaction Time Data. Psychol Rec. 2008 Jul;58(3):475–82.

63 Team RC. R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2019. Available from: http://www.R-project.org/.

64 Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using Lme4. J Stat Softw. 2015 Oct;67(1):1–48.

65 Fox J, Weisberg S. An R Companion to Applied Regression. London: Sage; 2018.

66 Aiken LS, West SG, Reno RR. Multiple Regression: Testing and Interpreting Interactions. London, UK: Sage; 1991.

67 Schielzeth H. Simple means to improve the interpretability of regression coefficients. Methods Ecol Evol. 2010 Jun;1(2):103–13.

68 Schielzeth H, Forstmeier W. Conclusions beyond support: Overconfident estimates in mixed models. Behav Ecol. 2009 Mar;20(2):416–20.

69 Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: Keep it maximal. J Mem Lang. 2013 Apr;68(3):255–78.

70 Baayen RH. Analyzing Linguistic Data: A Practical Introduction to Statistics Using R. Cambridge: Cambridge University Press; 2008. 369 p.

71 Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, et al. Generalized linear mixed models: A practical guide for ecology and evolution. Trends Ecol Evol. 2009 Mar;24(3):127–35.

72 Field AP. Discovering Statistics Using SPSS. 2nd ed. London: Sage; 2005. (Introducing Statistical Methods).

73 Tabachnick B, Fidell L. Using Multivariate Statistics. Boston, MA: Allyn and Bacon; 2001.

74 Quinn G, Keough M. Experimental Design and Data Analysis for Biologists. Cambridge: Cambridge University Press; 2002.

75 Zuur AF, Ieno EN, Elphick CS. A protocol for data exploration to avoid common statistical problems. Methods Ecol Evol. 2010 Feb;1(1):3–14.

76 Dobson AJ. An Introduction to Generalized Linear Models. Boca Raton: Chapman and Hall; 2002.

77 Forstmeier W, Schielzeth H. Cryptic multiple hypotheses testing in linear models: Overestimated effect sizes and the winner's curse. Behav Ecol Sociobiol. 2011 Jan;65(1):47–55.

78 Barton K. Package 'MuMIn': Multi-model inference [Internet]. R package version 1.43.17; 2020. Available from: https://CRAN.R-project.org/package=MuMIn.

79 Nakagawa S, Schielzeth H. Repeatability for Gaussian and non-Gaussian data: A practical guide for biologists. Biol Rev. 2010 Nov;85(4):935–56.

80 Stoffel MA, Nakagawa S, Schielzeth H. rptR: Repeatability estimation and variance decomposition by generalized linear mixed-effects models. Methods Ecol Evol. 2017 Nov;8(11):1639–44.

81 Green P, MacLeod CJ. SIMR: An R package for power analysis of generalized linear mixed models by simulation. Methods Ecol Evol. 2016 Apr;7(4):493–8.

82 Burnham KP, Anderson DR, Burnham KP. Model selection and multimodel inference: A practical information-theoretic approach. 2nd ed. New York, NY: Springer; 2002. 488 p.

83 Burnham KP, Anderson DR, Huyvaert KP. AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. Behav Ecol Sociobiol. 2011 Jan;65(1):23–35.

84 Mazerolle MJ. Package 'AICcmodavg': Model selection and multimodel inference based on (Q)AIC(c) [Internet]. R package version 2.2.2; 2019. Available from: https://CRAN.R-project.org/package=AICcmodavg.

85 Prescott MJ, Langermans JA, Ragan I. Applying the 3Rs to non-human primate research: Barriers and solutions. Drug Discov Today Dis Models. 2017 Mar;23:51–6.

86 Fuchs AF. Saccadic and smooth pursuit eye movements in the monkey. J Physiol. 1967 Aug;191(3):609–31.

87 Waitt C, Buchanan-Smith HM. Perceptual considerations in the use of colored photographic and video stimuli to study nonhuman primate behavior. Am J Primatol. 2006 Nov;68(11):1054–67.

88 Holmes A, Winston JS, Eimer M. The role of spatial frequency information for ERP components sensitive to faces and emotional facial expression. Cogn Brain Res. 2005 Oct;25(2):508–20.

89 Cronin KA, Bethell EJ, Jacobson SL, Egelkamp C, Hopper LM. Evaluating mood changes in response to anthropogenic noise with a response-slowing task in three species of zoo-housed primates. Anim Behav Cogn. 2018 Apr 30;5(2):209–21.

90 Plavcan JM. Sexual dimorphism in primate evolution. Am J Phys Anthropol. 2002 Jan;116(S33):25–53.

91 Angst W. Basic data and concepts on the social organization of *Macaca fascicularis*. In: Rosenblum LA, editor. Primate Behavior: Developments in Field and Laboratory Research. New York, NY: Academic Press; 1975. p. 325–88.

92 Bliss-Moreau E, Moadab G, Machado CJ. Monkeys preferentially process body information while viewing affective displays. Emotion. 2017 Aug;17(5):765–71.

93 Gronenschild EHBM, Smeets F, Vuurman EFPM, van Boxtel MPJ, Jolles J. The use of faces as stimuli in neuroimaging and psychological experiments: A procedure to standardize stimulus features. Behav Res Methods. 2009 Nov;41(4):1053–60.

94 Algom D, Chajut E, Lev S. A rational look at the emotional Stroop phenomenon: A generic slowdown, not a Stroop effect. J Exp Psychol Gen. 2004 Sep;133(3):323–38.

95 Grassini S, Railo H, Valli K, Revonsuo A, Koivisto M. Visual features and perceptual context modulate attention towards evolutionarily relevant threatening stimuli: Electrophysiology evidence. Emotion. 2019 Mar;19(2):348–64.

96 Mendl M. Performing under pressure: Stress and cognitive function. Appl Anim Behav Sci. 1999 Dec;65(3):221–44.

97 Bradley MM, Cuthbert BN, Lang PJ. Picture media and emotion: Effects of a sustained affective context. Psychophysiology. 1996 Nov;33(6):662–70.

98 Gray H, Thiele A, Rowe C. Using preferred fluids and different reward schedules to motivate rhesus macaques (*Macaca mulatta*) in cognitive tasks. Laboratory Animals. 2018 Oct 3;53(4):372–82.

99 Boissy A, Manteuffel G, Jensen MB, Moe RO, Spruijt B, Keeling LJ, et al. Assessment of positive emotions in animals to improve their welfare. Physiol Behav. 2007 Oct;92(3):375–97.

Figure legends

Fig. 1. Experimental time course of the dot-probe task. Panel (a) depicts the time course of two example trials of the dot-probe task. (1) Start button appeared within colored bar of the lower third of the touchscreen until touched or for five seconds. (2) An example of whole face stimuli appeared for 100 ms or 1000 ms. (3) The dot-probe replaced aggressive stimulus in the congruent position until touched or for 10 s. 0.25 ml of diluted grape juice was dispensed and a ding sound occurred if the dot-probe was touched correctly. Otherwise, a buzz sound occurred and a time penalty of 750 ms was added if the dot-probe was touched incorrectly (i.e., background touched) or did not touch the screen at all. (4) Each trial was followed by a blank screen for an inter-trial interval of 750 ms. (5) Start button reappeared until touched or for five seconds. (6) Example of scrambled stimuli (scrambled versions of stimuli in panel 2) presented for 100 ms or 1000 ms. (7) The dot-probe replaced the neutral stimulus in the incongruent position until touched or for 10 s. (8) 750 ms inter-trial interval. Panel (b) depicts the location of the dot-probe in relation to the whole face stimuli. The filled white circle illustrates the congruent condition, where the dot-probe replaces the aggressive stimulus. The dashed lined circle illustrates the incongruent condition, where the dot-probe replaces the neutral stimulus.

Fig. 2. Investigating if our dot-probe task can detect differences in monkeys' responses to dot-probe task parameters (congruency, trial type) during the baseline test session for 100 ms trials. (a) Mean reaction time per monkey to congruent and incongruent trials, connected by a thin black line. (b) Mean reaction time per monkey to scrambled or whole face trials, connected by a thin black line. Point area indicates the number of trials per condition, ranging from 28 to 35 trials. The Y-axes are scaled according to the transformed data. Model estimates are indicated by thick horizontal lines for each condition when all other predictors are at their mean (either dummy coded or z-transformed).

Fig. 3. Investigating if monkeys' responses to congruent and incongruent trials and attention bias changed on the day immediately after prolonged anesthesia (A + 1d) for 100 ms trials. Data shown were taken at baseline and on the day immediately after prolonged anesthesia (A + 1d). The dashed vertical line separates baseline data from A +1d data. (a) Mean reaction time per monkey for the interaction of congruency and test session, connected by a thin black line. (b) Mean attention bias score (mean response per monkey to incongruent trials subtracting that of congruent trials) per monkey for each test session (baseline, A + 1d). Positive values indicate a bias towards threatening content, whereas a negative value indicates a bias away from such content. Point area indicates the number of trials per condition for reaction time data and the number of trials used to calculate the attention bias score per test session, ranging from 20 to 68 trials. The Y-axes of reaction time plots

30

(figure a and c) are scaled according to the transformed data; values of the Y-axis of the attention bias score plot (figure b) are on the scale of ten to the minus five power due to score being calculated using the transformed response. Model estimates are indicated by thick horizontal lines for each condition when all other predictors are at their mean (either dummy coded or z-transformed).

Fig. 4. Investigating when monkeys' responses to congruent and incongruent trials returned to baseline levels following prolonged anesthesia for 100 ms trials. Data shown were taken at baseline, three- (A + 3d), seven- (A + 7d), and 14-days (A + 14d) after anesthesia. The dashed vertical line separates baseline data from data collected in the two weeks following anesthesia. Data taken on the one immediately following anesthesia (A + 1d) are not shown as they were not included in the model. Mean reaction time per monkey to congruent and incongruent trials per test session, connected by a thin black line. The final model indicates the test sessions occurring three-, seven-, and 14-days after anesthesia did not differ in their congruency pattern from baseline. Point area indicates the number of trials per condition, ranging from 23 to 35 trials. The Y-axes are scaled according to the transformed data. Model estimates are indicated by thick horizontal lines for each condition when all other predictors are at their mean (either dummy coded or z-transformed).